

Dual-decoder transformer for inverse design of optical multilayer films

Hua Wu, Shuhao Yuan, Bingcheng Li, Jia Peng, Haizheng Zhong and Yongyou Zhang

Citation: Wu H, Yuan SH, Li BC et al. Dual-decoder transformer for inverse design of optical multilayer films. *Intelligent Opto-Electronics* 2, 250018 (2026).

<https://doi.org/10.67704/ioe.2026.250018>

Received: 11 March 2026; Accepted: 4 March 2026; Published online: 25 March 2026

Related articles

OptoGPT: A foundation model for inverse design in optical multilayer thin film structures

Taigao Ma, Haozhu Wang, L. Jay Guo

Opto-Electronic Advances 2024, 7(7): 240062 doi: [10.29026/oea.2024.240062](https://doi.org/10.29026/oea.2024.240062)

Modelling-guided inverse design strategy for semitransparent perovskite photovoltaics with customized colors

Seok-Beom Seo, Rira Kang, Eun-Joo Lee, et al

Opto-Electronic Advances 2026, 9(2): 250218 doi: [10.29026/oea.2026.250218](https://doi.org/10.29026/oea.2026.250218)

Inverse design and realization of an optical cavity-based displacement transducer with arbitrary responses

Qianbo Lu, Qingxiong Xiao, Chengxiu Liu, et al

Opto-Electronic Advances 2023, 6(3): 220018 doi: [10.29026/oea.2023.220018](https://doi.org/10.29026/oea.2023.220018)

More related articles in Opto-Electronic Journals Group website 



Dual-decoder transformer for inverse design of optical multilayer films

Hua Wu¹, Shuhao Yuan¹, Bingcheng Li¹, Jia Peng², Haizheng Zhong² and Yongyou Zhang^{1*}

Abstract: Optical multilayer films (OMFs) are essential for modern photonics, yet their inverse design faces scalability and interpretability challenges due to high-dimensional parameters. Conventional optimization struggles with complexity, while existing machine-learning methods suffer from combinatorial explosion when jointly encoding materials and thicknesses. We present OptoFormer, a transformer network that addresses these issues via a dual-decoder design: one predicts material sequences and the other predicts thickness ones. Additionally, a material-compatible dictionary is introduced to ensure physical feasibility of the predicted OMFs. The model achieves high spectral prediction accuracy and autonomously learns physical correlations like structural inverse symmetry without explicit constraints. In simulation, we demonstrate a 36-layer omnidirectional reflector, high-Q optical cavities, and a hyperspectral light-coding array. OptoFormer provides a scalable and data-driven framework for the on-demand inverse design of advanced photonic devices.

Keywords: optical multilayer films; inverse design; dual-decoder transformer; deep learning

DOI: [10.67704/ioe.2026.250018](https://doi.org/10.67704/ioe.2026.250018)

Citation: Wu H, Yuan SH, Li BC et al. Dual-decoder transformer for inverse design of optical multilayer films. *Intell Opto-Electron* 2, 250018 (2026).

1 Introduction

Optical multilayer films (OMFs) are foundational to optical engineering, enabling critical components ranging from narrow-band filters to high-Q cavities^{1,2}. Configuring OMFs for specific spectral responses is a classic inverse problem, requiring the precise optimization of both material sequences and thickness ones^{3,4}. However, as modern applications demand increasingly complex OMFs with dozens of layers, the resulting high-dimensional parameter space presents scalability challenges^{5–7}. This complexity mirrors challenges in broader frontier fields like metamaterials^{8–20}. Conventional design methods for OMFs, rooted in iterative optimization, often falter in this regime^{21,22}. While global optimization techniques like genetic algorithms²³ and particle swarm optimization²⁴ work well in low-dimensional spaces, their computational cost becomes prohibitive for multi-layer structures, where they frequently converge to local optima²⁵.

Machine learning (ML) has emerged as a transformative paradigm for photonic inverse design, leveraging generative models^{26–28} and reinforcement learning to capture complex structure–property correlations^{29–34}. These data-driven frameworks have been widely validated across diverse nanophotonic applications, spanning metasurface absorbers^{35–38}, structural color generation³⁹, grating couplers⁴⁰, and topological photonics^{12,16}, effectively bridging the gap between structural design parameters and optical responses^{41–48}. For example, the neutron framework has been developed to realize material-aware photonic design⁴⁹, while proximal policy optimization has been adopted to advance automated optical inverse design^{50,51}. In parallel, reinforcement learning schemes such as deep Q-learning have been successfully employed to optimize the performance of optical coatings⁵². Despite the remarkable successes of these approaches, addressing the coupled discrete/continuous nature of OMFs with dynamically variable layer numbers remains a critical open challenge. Recently, the sequence-based foundational

Received: 11 March 2026

Accepted: 4 March 2026

Published online: 25 March 2026

¹School of Physics, Beijing Institute of Technology, Beijing 100081, China; ²MIIT Key Laboratory for Low-Dimensional Quantum Structure and Devices, School of Materials Science & Engineering, Beijing Institute of Technology, Beijing 100081, China.

*Correspondence: YY Zhang, E-mail: yyzhang@bit.edu.cn

model OptoGPT⁴ introduced a new language-inspired paradigm to tackle this challenge. However, it has a natural constraint in jointly encoding material-thickness composite tokens, since this encoding strategy implies a rapid growth of dictionary size when scaling to complex structures.

To address this specific scalability issue, we present OptoFormer, a transformer architecture that extends the sequence-based approach via a dual-decoder design that decouples the prediction of materials and thicknesses. One decoder exclusively predicts materials, while the other predicts thicknesses, and the two decoders execute their prediction tasks in an alternating manner. This decoupling enables OptoFormer to efficiently explore an expansive design space ($\sim 10^{101}$) spanning OMFs with up to 36 layers. Furthermore, a material-compatible dictionary is integrated into the decoding process to ensure the physical feasibility of the predicted structures. Simulation results demonstrate that OptoFormer maintains high spectral accuracy ($R^2 > 0.8$ for over 80% of samples up to $N_{\text{Layer}} = 30$) while autonomously discovering physical correlations, such as structural inverse symmetry. We demonstrate its capability by generating a 36-layer omnidirectional reflector, ultra-high-Q optical cavities ($Q \approx 1.2 \times 10^6$), and hyperspectral coding arrays in a one-shot inference mode. These results position OptoFormer as a robust platform for data-driven design, offering a scalable solution to the challenges of high-dimensional OMF exploration.

2 Methods

2.1 Structure serialization

OMFs are characterized by material and thickness sequences (see Fig. 1(a)), mathematically defined as

$$\mathbf{M} = [M_1, M_2, \dots, M_{N_{\text{Layer}}}], \mathbf{h} = [h_1, h_2, \dots, h_{N_{\text{Layer}}}],$$

where M_n and h_n denote the material and thickness value of the n -th layer, respectively. Unlike OptGPT, which serializes OMF structures using "material-thickness" token pairs, our approach encodes an N_{Layer} -layer film into two independent sequences of material and thickness tokens. Each sequence contains $N_{\text{Layer}} + 2$ tokens, including start (0) and end ($N_{\text{Material}} + 1$ for materials, $N_{\text{Thickness}} + 1$ for thicknesses) markers, to standardize predictions across diverse structures.

The simulation setup involves 8 materials ($N_{\text{Material}} = 8$: Al_2O_3 , CaF_2 , HfO_2 , MgF_2 , MgO , SiO_2 , TiO_2 , ZrO_2) and 81 thickness values ($N_{\text{Thickness}} = 81$: 20–420 nm in 5-nm increments). Material (thickness) indices span from 1 to N_{Material} (1 to $N_{\text{Thickness}}$), with start and end markers assigned as 0 and $N_{\text{Material}} + 1$ (0 and $N_{\text{Thickness}} + 1$), respectively. Tokenization of materials and thicknesses implies the OMF design is a classification task rather than a regression one. Transformer-based architectures excel in classification tasks, making this tokenization strategy well-suited for our model. The wavelength dependences of the refractive indices for all used materials are provided in Section 1 of the supplementary

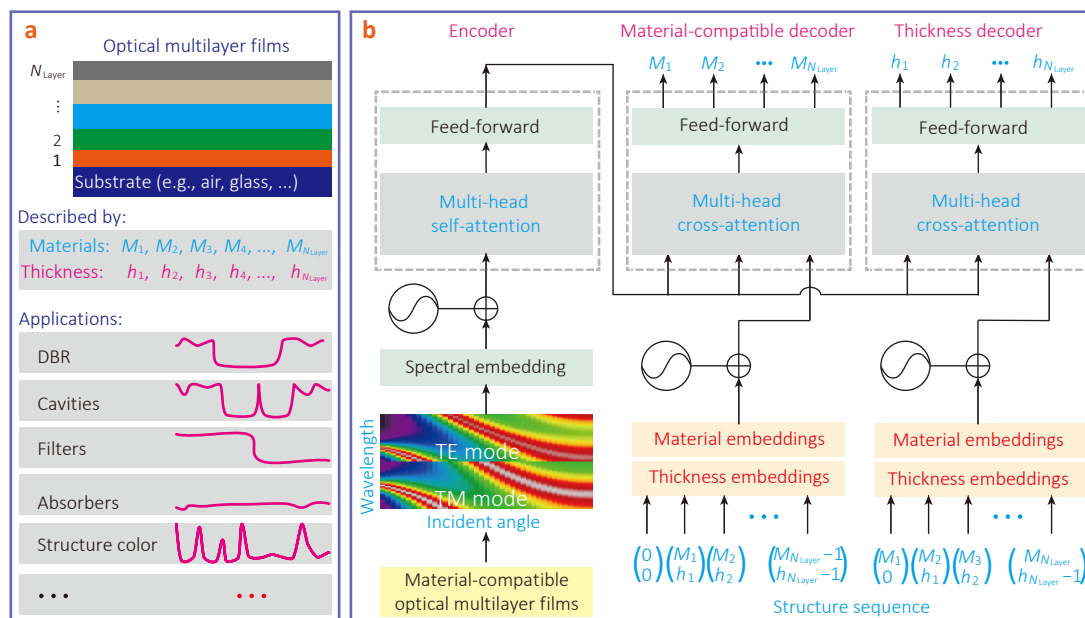


Fig. 1 | | Schematic of OptoFormer for designing OMFs. (a) Diagram of an OMF structure deposited on a customizable substrate (e.g., air, glass, et al.) and its potential applications. (b) Architecture of OptoFormer, comprising an encoder, material-compatible decoder, and thickness decoder. The encoder takes vertically concatenated TE and TM transmission spectra as input. The material decoder inputs the right-shifted sequence of previous material-thickness pairs (indices 0 to $N_{\text{Layer}} - 1$) to predict the material sequence. In contrast, the thickness decoder inputs the unshifted material sequence (indices 1 to N_{Layer}) combined with the right-shifted thickness sequence. This asymmetric input configuration allows the thickness decoder to condition on the current material token M_n to predict the corresponding thickness h_n , enforcing the $M_n \rightarrow h_n$ dependency.

information (SI). With N_{Layer} spanning from 2 to 36, the structure space expands to $(8 \times 81)^{36} \sim 10^{101}$, showcasing the vast design landscape that OptoFormer aims to explore under predefined spectral constraints.

2.2 Condition encoder

Unlike language models that take text as input, OptoFormer utilizes optical spectra as input. Since the material dictionary excludes absorptive materials, OptoFormer focuses on transmission spectra and does not consider reflection spectra. Additionally, OptoFormer can be directly extended to scenarios involving absorptive materials without architectural modifications, i.e., simply by integrating absorption or reflection spectra into the input spectra.

The input spectra in Fig. 1(b) are derived from material-compatible OMFs. The compatibility of the eight candidate materials is summarized in Table 1. A checkmark denotes experimentally validated compatibility between a given pair of materials, whereas a cross indicates the lack of confirmed compatibility. To facilitate subsequent coding implementations, the material compatibility is also organized into a dictionary, with material indices designated as keys and the corresponding indices of compatible materials assigned as values. The input spectra vertically concatenate the TE and TM transmission spectra, where the vertical axis represents wavelength and the horizontal axis represents incident angle. The wavelength ranges from 380 nm to 760 nm at 1 nm intervals, while the incident angle spans 0° to 80° at 2° intervals. Each input token vector incorporates both TE and TM transmission spectra, with the token length of the incident angles being 41 in this work. After applying positional encoding to the input spectra, the linear transformation layer (spectral embedding) serves as a critical bridge by projecting the position-integrated spectral data into the feature dimension space required by the Transformer model (d_{model}). Subsequently, the encoder processes the spectral embeddings through a stack of multi-head self-attention blocks interleaved with feed-forward networks, generating

contextualized codings that capture both local and global spectral dependencies. Collectively, the encoder projects the polarized angle-dependent transmission spectra into a latent space. These contextualized codings are then fed into the material decoder to predict materials and into the thickness decoder to predict thickness values (see Fig.1(b)).

2.3 Dual decoders

The primary objective of OMF prediction is to infer the material and thickness values of individual layers from their spectral reflection/transmission spectra. A critical challenge in current ML-based approaches lies in their inability to handle OMFs with variable layer counts N_{Layer} , such as three-layer structural color films or six-layer solar metamaterial absorbers³⁵. To address this limitation, the OptoGPT framework was introduced recently, leveraging large language model architectures for OMF inverse design⁴. OptoGPT encodes OMF structures into sequences of "material-thickness" pair tokens, transforming the inverse design task into discrete prediction of token indices. This serialization scheme enables generalization across OMFs with heterogeneous layer counts and material combinations. However, as mentioned above, the merged dictionary size scales multiplicatively with the cardinality of material and thickness sub-dictionaries. For example, if independent material and thickness dictionaries contain 10 and 100 entries, their Cartesian product yields a 1000-element "material-thickness" dictionary. To circumvent this combinatorial explosion, OptoFormer decouples the prediction task into two independent latent spaces. This decoupling introduces the primary computational overhead compared to single-decoder models such as OptoGPT⁴, as OptoFormer inherently involves two distinct decoding steps.

Material and thickness embeddings are fused into a unified feature vector, and the two decoders execute their prediction tasks in an alternating manner, as illustrated in Fig. 2(a). This alternating strategy leads to a key difference in merging embeddings for two decoders. Specifically, the material decoder applies synchronized right-shifting to both material

Table 1 | Compatibility of the eight candidate materials. A checkmark (✓) signifies experimentally validated compatibility between a given pair of materials, whereas a cross (×) indicates the lack of confirmed compatibility. The diagonal symbol “—” refers to the same material. The material compatibility is also organized into a dictionary structure, with material indices designated as keys and the corresponding indices of compatible materials assigned as values.

	1: Al ₂ O ₃	2: CaF ₂	3: HfO ₂	4: MgF ₂	5: MgO	6: SiO ₂	7: TiO ₂	8: ZrO ₂	Material-compatible dictionary
1: Al ₂ O ₃	—	×	×	✓ ⁵³	✓ ⁵⁴	✓ ⁵⁵	✓ ⁵⁶	✓ ⁵⁶	= {‘1’: [4, 5, 6, 7, 8]}
2: CaF ₂		—	×	✓ ⁵⁷	✓ ⁵⁸	✓ ⁵⁹	✓ ⁶⁰	×	‘2’: [4, 5, 6, 7]
3: HfO ₂			—	✓ ⁶¹	×	✓ ⁶²	✓ ⁶³	×	‘3’: [4, 6, 7]
4: MgF ₂				—	×	✓ ⁶⁴	✓ ⁶⁵	✓ ⁶⁶	‘4’: [1, 2, 3, 6, 7, 8]
5: MgO					—	✓ ⁶⁷	×	✓ ⁵⁶	‘5’: [1, 2, 6, 8]
6: SiO ₂						—	✓ ⁶⁸	✓ ⁶²	‘6’: [1, 2, 3, 4, 5, 7, 8]
7: TiO ₂							—	✓ ⁵⁶	‘7’: [1, 2, 3, 4, 6, 8]
8: ZrO ₂								—	‘8’: [1, 4, 5, 6, 7]}

and thickness token sequences (indices 0 to $n - 1$) to predict the next material M_n . In contrast, the thickness decoder inputs the unshifted material sequence (indices 1 to n) combined with the right-shifted thickness sequence (indices 0 to $n - 1$). This asymmetric configuration allows the thickness decoder to condition on the current material token M_n to predict the corresponding thickness h_n , enforcing the physical dependency between the chosen material and its layer thickness (see Fig. 2(a)). Note that when the material decoder outputs the end token for the material sequence, both the material and thickness decoders cease their predictions simultaneously. Such a mechanism ensures that the material and thickness sequences are of identical length. This alternating prediction embodies physical intuition: materials, distinguished by their wavelength-dependent dielectric constants, are inherently more classifiable than thicknesses, especially when thickness intervals are small. Moreover, each material inherently requires a corresponding thickness value, reflecting the deterministic relationship between materials and thicknesses in OMFs. In addition, the material-compatible dictionary in Table 1 is integrated into the material decoder by only considering compatible predicted materials and scoring these compatible ones to ensure the practice of the predicted OMF structures. OptoFormer's architecture is provided in Section 2 of the SI.

To train OptoFormer, we generate transmission spectra for OMF structures with layer counts N_{layer} ranging from 2 to 36. Spectral generation is implemented via a stochastic sampling strategy to cover the expansive OMF design space, see details provided in Section 3 of the SI. The transmission spectra are calculated using the transfer matrix method (TMM) in all simulations. Without loss of generality, air was adopted as the substrate for all generated samples in this work to demonstrate OptoFormer's performance. Note that the OptoFormer framework is substrate-agnostic and can be readily applied to arbitrary substrate materials (e.g., glass, silicon) by simply adjusting the boundary conditions in

TMM calculations, with no architectural modifications required for the model itself.

3 Results

3.1 Tokens visualization

To characterize the performance of the trained OptoFormer, we utilize t-distributed stochastic neighbor embedding (t-SNE) and uniform manifold approximation and projection (UMAP) to visualize the latent embeddings of material and thickness tokens. These high-dimensional embeddings are projected into two-dimensional space for visualization, as shown in Fig. 2(b).

Material tokens exhibit clear separation with distinguishable inter-token distances, as illustrated in the left-column panels of Fig. 2(b). The inter-token distances are consistent with physical intuition: materials with similar refractive indices tend to cluster closely together (e.g., MgF_2 and SiO_2 form one cluster, while TiO_2 and ZrO_2 form another). This behavior also holds for thickness tokens, as shown in the middle-column panels of Fig. 2(b), where thickness tokens display a manifold structure that follows the monotonic increase of thickness values. When the embeddings of material and thickness tokens are combined into composite (M_n, h_n) embeddings, tokens corresponding to the same material form compact and concentrated clusters, as depicted in the right-column panels of Fig. 2(b). This observation is in contrast to the linear distribution observed in OptoGPT⁴. Accordingly, OptoFormer can not only learn the intrinsic physical representations of OMFs but also achieve more effective separation among all tokens.

3.2 OptoFormer's performance

To evaluate OptoFormer's performance, we assess its accuracy

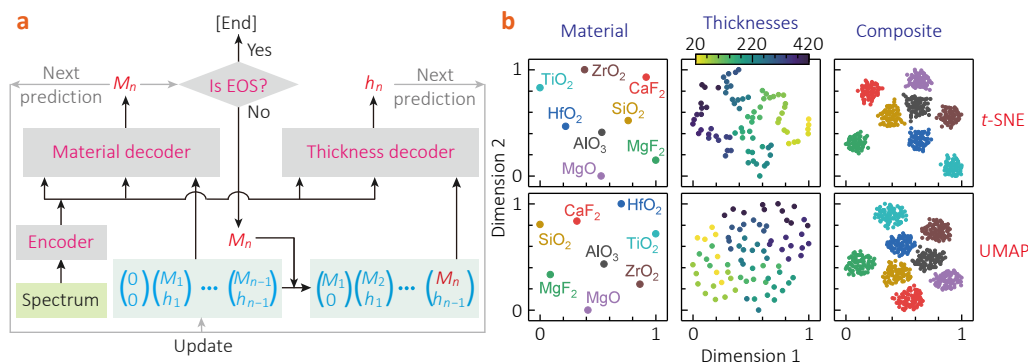


Fig. 2 | (a) Dual-decoder inference process. At step n , the material decoder first predicts the material token M_n . This token is immediately checked against the End-of-Sequence (EOS) marker. If the token is valid, M_n serves as a conditional input to the thickness decoder for predicting the corresponding thickness h_n . The generated pair (M_n, h_n) is then fed into the sequence history for the next iteration. (b). Visualization of the learned latent embeddings using t-SNE (top row) and UMAP (bottom row). Panels from left to right correspond to material (M_n), thickness (h_n), and their combined (M_n, h_n) tokens. The colors in the right-column panels denote the same materials as in the left-column panels.

in predicting OMFs with layer numbers N_{Layer} ranging from 2 to 36. The evaluation process is as follows. i) For each N_{Layer} , generating 10,000 OMF test samples (targeted structures) and corresponding transmission spectra (targeted spectra); ii) Feeding these targeted spectra into OptoFormer to predict the OMF structures (predicted structures); iii) Calculating the transmission spectra of these predicted structures (predicted spectra) by transfer matrix theory; iv) Calculating the coefficient of determination (R^2) between the targeted spectra and the predicted ones. The performance metric is defined as the proportion of samples with $R^2 > \chi$, where χ is a predefined threshold (e.g., 0.9). This proportion is plotted as a function of N_{Layer} in Fig. 3(a) for four cases with $\chi = 0.8, 0.85, 0.90$, and 0.95 .

The proportion of high- R^2 samples decreases as N_{Layer} increases. For all four thresholds, the proportion of qualified samples approaches 100% when $N_{\text{Layer}} \lesssim 10$. Specifically, when $N_{\text{Layer}} \lesssim 20$, 90% of samples exhibit $R^2 > 0.95$, demonstrating the model's strong predictive accuracy for moderately deep structures. As N_{Layer} increases to 30, about 80% of samples maintain $R^2 > 0.8$, and this proportion further declines to approximately 35% for $N_{\text{Layer}} = 36$. This degradation is attributed to the exponentially increasing prediction complexity arising from the combinatorial explosion of the structure space (up to 10^{101} for 36 layers), which challenges the model's ability to capture long-range structural dependencies. Statistical errors, represented by the background shadows in Fig. 3(a), remain below $\sim 3.1\%$ for $N_{\text{Layer}} \lesssim 25$ and reach a maximum of $\sim 6.2\%$ at $N_{\text{Layer}} = 32$, indicating consistent performance within the model's optimal layer range. As a result, OptoFormer delivers robust prediction accuracy ($R^2 > 0.8$ for over 80% of samples) when $N_{\text{Layer}} \lesssim 30$, making it a viable solution for practical OMF inverse design tasks within this regime. If one wishes to enable OptoFormer to reliably cover OMFs with more than 30 layers, scaling up the model architecture would be a

viable solution.

A large R^2 value indicates a small mean absolute error (MAE) between the predicted and target spectra. Specifically, the MAE values increase with an increase in N_{Layer} which is an expected trend for the prediction of multilayer structures. When considering all tested OMFs with $N_{\text{Layer}} \leq 20$, OptoFormer achieves a slightly lower MAE (0.0255) compared to OptoGPT (0.0258)⁴. For high-quality predictions (i.e., those with $R^2 > 0.8$), the MAE of OptoFormer is further reduced to 0.0165. More detailed analyses about the MAE are provided in Section 4 of the SI.

Figure 3(b,c) present three representative cases to illustrate the prediction performance of OptoFormer. Specifically, Fig. 3(b) compares the targeted and predicted transmission spectra, while Fig. 3(c) shows the corresponding OMF structures as ordered material sequences, with thickness values enclosed in parentheses. All predicted OMF structures exhibit material compatibility. For $N_{\text{Layer}} = 2$, the predicted spectra and structure perfectly align with the targets, confirming the model's capability to learn the spectrum-to-structure mapping. For $N_{\text{Layer}} = 12$, although the predicted spectra align well with the targets, the predicted structure is inversely ordered relative to the target. This observation indicates that OptoFormer has learned to leverage structural inverse symmetry. Significantly, this symmetry was not explicitly encoded during training, implying the model infers fundamental physics-driven correlations between spectral signatures and structural configurations. For such cases, the predicted OMFs are structurally identical to the target ones, leading to perfect matching between the predicted and target spectra.

For $N_{\text{Layer}} = 24$, the predicted structure remains inversely ordered relative to the target, except for slight deviations in the thickness values of the layers highlighted in magenta and blue in Fig. 3(c). Despite this, the primary features of the predicted spectra still match those of the targeted spectra

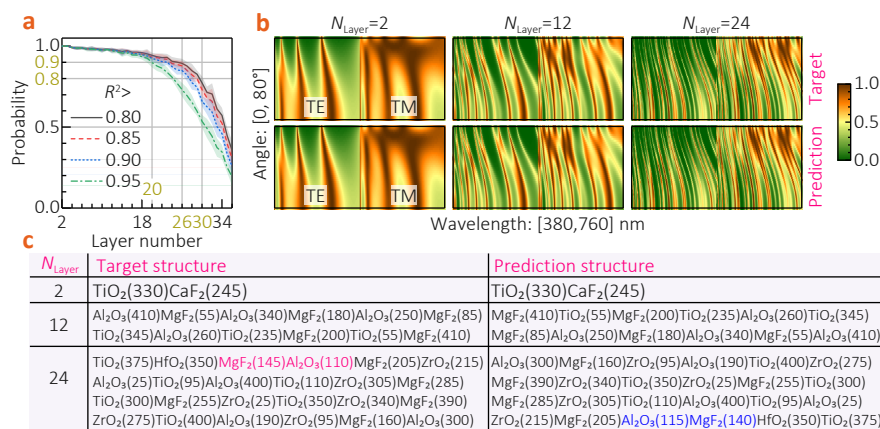


Fig. 3 | Performance of OptoFormer. (a) Proportions of OMF samples with $R^2 > \chi$ ($\chi = 0.8, 0.85, 0.90$, and 0.95) as a function of N_{Layer} . (b) Targeted and predicted transmission spectra for three representative cases. (c) Corresponding targeted and predicted OMF structures for the cases in (b), with thickness values (in nm) in parentheses.

(Fig. 3(b)); however, the minor thickness deviations induce a small shift (several nanometers) of the peaks in the predicted spectra relative to the target ones. To clearly demonstrate this effect, the absolute error spectra between the predicted and target cases are provided in Section 5 of the SI. While the maximum absolute error reaches ~ 0.52 , the MAE remains as low as ~ 0.01 . Quantitatively, the predicted thickness of the Al_2O_3 layer (highlighted in blue) is 115 nm, slightly lower than the target value (highlighted in magenta), whereas the predicted thickness of MgF_2 is 140 nm, slightly higher than the target. This result reveals that a compensation mechanism is at play when OptoFormer predicts OMFs. Consequently, the model tunes the overall optical response to match the target spectra. This behavior highlights the model's ability to learn implicit physical correlations between structural modifications and spectral outcomes, balancing local thickness deviations against global optical performance.

3.3 Omnidirectional reflectors

A trained OptoFormer can predict desired OMFs from transmission spectra. This subsection demonstrates OptoFormer's capabilities by using the prediction of omnidirectional reflectors (ORs) in visible regime. To achieve this, the input transmission spectra are set as a zero tensor. It is crucial to note that randomly generated or researcher-specified transmission spectra generally lack corresponding OMFs in physics, although each OMF has unique transmission spectra. Thus, an ideal OR in visible regime may not exist. To address this, small random values (i.e., random

perturbations with an amplitude of 0.05) are added to the input transmission spectra, ensuring they are not strictly zero. Among 200 predictions, the optimal result is selected as the final predicted structure, with performance evaluated by the R^2 between the predicted transmission spectra and the zero-value one. The OR's transmission spectra for TE and TM modes are shown in Fig. 4(a), and the predicted 36-layer reflector structure, divided into seven sections for analysis, is depicted in Fig. 4(b), with thickness values in parentheses. OptoFormer employs merely two materials: CaF_2 , which exhibits the second-lowest refractive index in the material dictionary, and TiO_2 , which boasts the highest value. This selection indicates that the model inherently prioritizes materials with large refractive index contrasts. Note that CaF_2 is compatible with TiO_2 ⁶⁰.

The first six layers form a distributed Bragg reflector (DBR) with only one thickness deviation: the fifth-layer CaF_2 is 95 nm instead of 100 nm (line I in Fig. 4(b)). This aligns with the physical intuition that DBR structures are favored for omnidirectional reflection. Using the average refractive indices of CaF_2 ($n_{\text{CaF}_2} = 1.435$) and TiO_2 ($n_{\text{TiO}_2} = 2.981$) in visible regime, the reflection center wavelength of the DBR is approximately $2 \times (100 \times 1.435 + 45 \times 2.981) \approx 555$ nm, close to the center of the visible band (380–760 nm). This DBR exhibits low transmission in the green band but high transmission in the blue and red bands (see $N_{\text{Layer}} = 6$ spectra in Fig. 4(a)). The second DBR (line II in Fig. 4(b)) with a center wavelength of $2 \times (55 \times 1.435 + 45 \times 2.981) \approx 426$ nm is used to attenuate blue-band transmission (see $N_{\text{Layer}} = 10$ spectra in Fig. 4(a)). However, a prominent transmission peak at ~ 600 nm arises from the

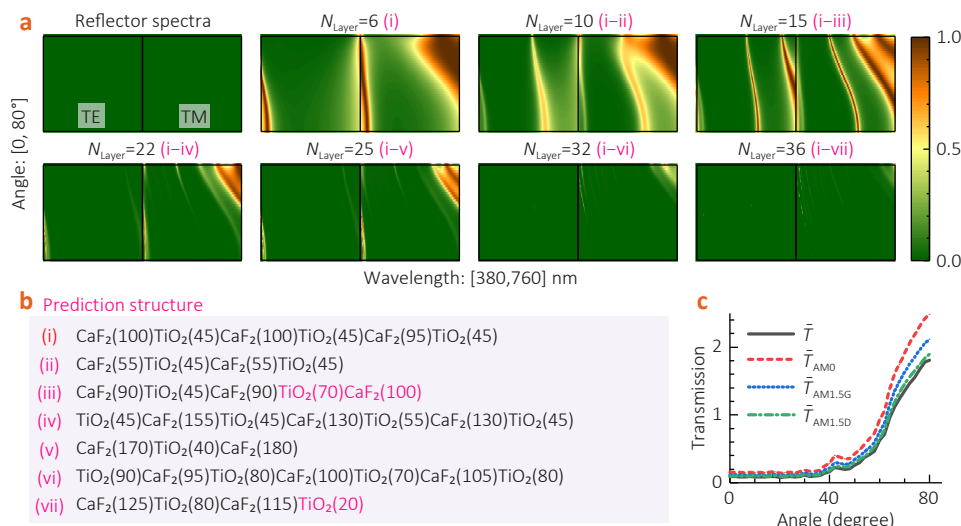


Fig. 4 | Omnidirectional reflectors (ORs) predicted by OptoFormer. (a) Transmission spectra of the OR (targeted reflector spectra) and OMFs using only the first N_{Layer} layers of the total 36-layer predicted structure. The Roman numerals in parentheses denote the layer sequences adopted, as listed in (b). (b) Predicted structure of the OR: a 36-layer OMF divided into seven sections, with thickness values (in nm) shown in parentheses. (c) Variation of the mean transmissions of the predicted OR in (b) with incident angle across the visible spectrum (380–760 nm). $\bar{T}$ represents the geometric mean transmission, while the other three metrics correspond to three types of solar spectra (AM0, AM1.5G, AM1.5D).

combined DBR structure. To address this, OptoFormer increases the CaF_2 thickness from 55 nm to 90 nm in layers 11–13 (line III in Fig. 4(b)). Further thickness adjustments in layers 14–15 result in a significantly narrow 600 nm peak (see $N_{\text{Layer}} = 15$ spectra in Fig. 4(a)).

With layers 16–22 (line IV in Fig. 4(b)), the model targets suppression of long-wavelength transmission at large incident angles, which uses a quasi-DBR with a ~ 700 nm center wavelength (see $N_{\text{Layer}} = 22$ spectra in Fig. 4(a)). Anticipating blueshift with increasing angles, CaF_2 thicknesses are increased to >170 nm in layers 23–25 (line V in Fig. 4(b)), forming a quasi-DBR with a ~ 740 nm center wavelength. Given that 740 nm remains below the visible band's upper limit (760 nm), the model adjusts its strategy in layers 26–32 by increasing TiO_2 thicknesses (line VI in Fig. 4(b)). The corresponding quasi-DBR's center wavelength is ~ 760 nm, thereby strongly suppressing long-wavelength transmission at large angles (see $N_{\text{Layer}} = 32$ spectra in Fig. 4(a)). Layers 33–35 (line VII in Fig. 4(b)) further enhance this suppression effect with the final quasi-DBR whose center wavelength is ~ 820 nm. The $N_{\text{Layer}} = 36$ spectra in Fig. 4(a) confirm an effective OR obviously.

The predicted OR can be further evaluated by examining the mean transmission as a function of incident angle. The mean transmission, defined as

$$\bar{T}(\theta) = \frac{1}{380} \int_{380}^{760} T(\lambda, \theta) d\lambda,$$

is plotted against the incident angle in Fig. 4(c), where $T(\lambda, \theta)$ denotes the transmission at wavelength λ and incident angle θ . The quantities $\bar{T}_{\text{AM0}}$, $\bar{T}_{\text{AM1.5G}}$, and $\bar{T}_{\text{AM1.5D}}$ represent the mean transmissions of the AM0, AM1.5G, and AM1.5D solar spectra through the predicted OR, respectively. All four metrics in Fig. 4(c) exhibit low transmission at small incident angles, with values increasing to only 2% at 80° . As a result, OptoFormer has successfully generated the OR with high omnidirectional reflection performance across the visible regime.

3.4 Optical cavities

A direct application of the predicted OR is in optical cavity design. Here, we utilize it to demonstrate OptoFormer's capabilities by designing optical cavities. The preceding analysis of OR prediction reveals that any number of preceding layers among the total 36 can form a functional OR. OR performance improves with increasing N_{Layer} (spectra in Fig. 4(a)). Figure 5 illustrates cavity structures composed of these ORs (first row) and their corresponding normal-incidence transmission spectra (second row). The cavity layer is CaF_2 , sandwiched between two identical ORs (left and right), as shown in the schematic diagrams of Fig. 5. The right OR is a mirror image of the left one. Roman numerals denote the layer sets used in the ORs, as specified in Fig. 4(b). For brevity, these ORs are denoted as OR -

(Roman numerals).

As the cavity height h increases from 200 nm to 500 nm in 10 nm increments, the transmission peaks exhibit the expected redshift (Fig. 5(a)). For the cavity using OR-(I) (six layers, line I in Fig. 4(b)), the two marked peaks have full widths at half maximum (FWHMs) of approximately 2.8 nm and 10 nm, corresponding to cavity quality factors $Q \approx 170$ and 65, respectively. When OR-(I–III) is employed (15 layers, lines I–III in Fig. 4(b)), the transmission peaks become narrow significantly (Fig. 5(b)), with the minimum FWHM among marked peaks being ~ 0.017 nm and the corresponding $Q \approx 2.8 \times 10^4$. For OR-(I–V) (25 layers, lines I–V in Fig. 4(b)), the FWHM reaches 3.9×10^{-3} nm, yielding $Q \approx 1.4 \times 10^5$ (Fig. 5(c)). Using OR-(I–VII) (36 layers, lines I–VII in Fig. 4(b)), the FWHM can be as low as 5.0×10^{-4} nm, corresponding to $Q \approx 1.2 \times 10^6$ (Fig. 5(d)). Such a high- Q cavity demonstrates that OptoFormer exhibits the capability to design optical structures for device applications.

3.5 Hyperspectral imaging

The optical cavities designed above hold significant potential for applications in hyperspectral imaging, a technology critical for resolving fine spectral details in fields ranging from remote sensing to biomedical diagnostics. As illustrated in Fig. 5(e), the hyperspectral imaging scheme operates by first transmitting incident light through a coding layer, composed of the 31 optical cavities designed in Fig. 5(a). That is, the optical coding layer encodes the spectral information into a 31-element vector. This encoded signal is then detected by a CCD camera, serving as the low-dimensional measurement basis for subsequent spectral reconstruction. The underlying principle leverages compressed sensing theory, where the coding array acts as a physical realization of a measurement matrix, enabling efficient spectral information acquisition by exploiting the inherent sparsity of real-world spectra^{45,69–71}. Conventional spectral reconstruction methods, including least squares, ridge regression, Lasso regression, compressed sensing reconstruction, and sparse dictionary learning, rely on linear or weakly nonlinear models to invert the coding process. While these techniques excel in noise-free or moderately sparse scenarios, they often falter when faced with high-frequency noise or complex spectral structures (e.g., overlapping absorption bands or non-Gaussian line shapes). Such limitations arise from their reliance on handcrafted features or restrictive sparsity assumptions. Accordingly, the ML-based reconstruction framework depicted in Fig. 5(f), featuring a fully-connected layer (FCL) encoder-decoder pair, is designed to learn end-to-end mapping between incident-light spectra and their coded measurements. The architecture is provided in Section 6 of the SI.

To assess the framework's performance, we generated a diverse dataset of synthetic spectral samples, each

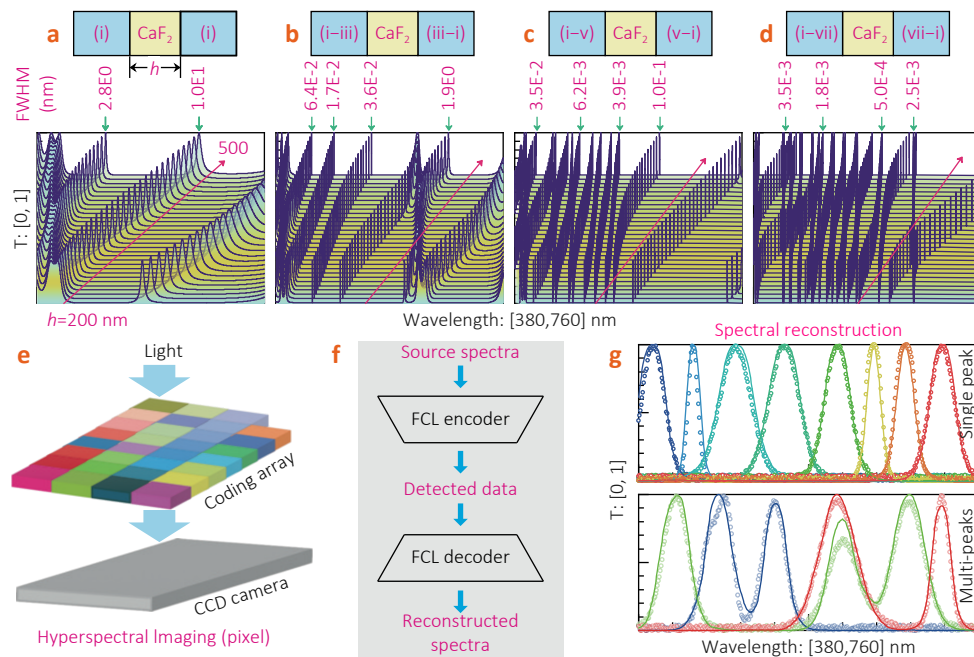


Fig. 5 | Optical cavities and their potential application in hyperspectral imaging. (a-d) Optical cavity designs and their normal-incidence transmission spectra, with the cavity layer (CaF_2) thickness h varying from 200 nm to 500 nm in 10 nm increments. The ORs on both sides of the cavities, composed of the OMF sequences detailed in Fig. 4(b), exhibit mirror symmetry. (e) Hyperspectral imaging schematic. The incident light is first transmitted through a coding array and then detected by a CCD camera. Using the detected data, the spectrum of the incident light can be reconstructed. The coding array shown here is composed of 31 optical cavities as designed in (a). (f) Spectral reconstruction neural network framework applied in (e). The incident-light spectra (from 380 nm to 760 nm at 1 nm interval) are fed into a fully-connected-layer (FCL) encoder to predict the detected data (31 components), and the FCL decoder uses the detected data to reconstruct the incident-light spectra. (g) Comparison between reconstructed spectra (lines) and source spectra (dots) for single-peak (upper panel) and multi-peak (lower panel) spectra.

contaminated with 4% additive white Gaussian noise. The samples consist of 1–3 Gaussian spectral components with full-width at half-maximum bandwidths ranging from 5 nm to 20 nm, emulating typical sparsely structured spectral features in hyperspectral imaging. This design ensures the dataset captures both simple (single-peak) and complex (multi-peak, overlapping) spectral scenarios, rigorously testing the model's generalization capability across diverse signal topologies. As illustrated in Fig. 5(g), the reconstruction results exhibit high-fidelity spectral recovery for both single-peak (upper panel) and multi-peak (lower panel) configurations. The solid lines (reconstructed spectra) show close alignment with the circular dots (incident spectra with 4% white noise). Due to the inherently higher sparsity of single-peak spectra, their spectral overlap with reconstructed counterparts achieves remarkable accuracy. While multi-peak spectra pose greater reconstruction challenges due to spectral complexity and potential interference, the model also effectively captures their dominant features, as evident in the lower panel of Fig. 5(g). The high reconstruction fidelity indicates that the model efficiently suppresses measurement noise, showcasing the effectiveness of OptoFormer in designing optical devices.

To further verify the model's generalization capability

across diverse spectral regimes beyond the resonant structures presented here, we also provide design examples for broadband anti-reflection coatings, sharp edge filters, and narrowband transmission filters in Section 7 of the SI.

4 Discussion and conclusion

OptoFormer demonstrates high performance in predicting OMFs with up to 36 layers, showcasing its applicability to complex photonic structures such as omnidirectional reflectors and high-Q optical cavities. OptoFormer's decoding strategy effectively addresses the combinatorial explosion issue in prior ML approaches, making it scalable to large structure spaces ($\sim 10^{101}$) while maintaining prediction accuracy. While OptoFormer learns thickness distributions, predicting consecutive thickness values with high precision remains challenging, as evidenced by the compensatory thickness adjustments in prediction process. This limitation can be mitigated by incorporating hierarchical attention mechanisms to enhance long-term correlation, increasing the sample ratio for extremely deep OMFs, or enlarging OptoFormer's architecture. Additionally, the OptoFormer's architecture is extendable to include reflection/absorption data, though its reliance on transmission spectra in our

simulations (excluding absorption) limits its applicability to lossy materials. As demonstrated in⁴, task-specific conditions can also be incorporated into the inference procedure to enable conditional generation of OMFs.

To summarize, we present OptoFormer, a transformer-based architecture for the inverse design of optical multilayer films. By decoupling material and thickness predictions, the model overcomes the combinatorial complexity of prior methods, enabling efficient exploration of OMFs with up to 36 layers. Material compatibility is integrated into training data generation and material decoder to guarantee the practical application of the predicted structures. Simulation results demonstrate high prediction accuracy and the ability to infer physical correlations (e.g., structural inverse symmetry) without explicit physical constraints. As proof-of-concept, we design a 36-layer omnidirectional reflector with mean transmission < 2% across the visible regime, an optical cavity with quality factors reaching 1.2×10^6 , and a light-coding array in hyperspectral imaging. They highlight OptoFormer's potential as a transformative tool for generating high-fidelity photonic devices directly from spectral targets, providing high-fidelity OMF designs that are ready for fabrication within the tolerance range of lithography processes.

References

- Pan ZY, Pan XM. Deep learning and adjoint method accelerated inverse design in photonics: a review. *Photonics* **10**, 852 (2023).
- MacLeod HA. *Thin-Film Optical Filters* 4th ed (CRC Press, Boca Raton, 2010).
- Ma TG, Ma MQ, Guo LJ. Optical multilayer thin film structure inverse design: from optimization to deep learning. *iScience* **28**, 112222 (2025).
- Ma TG, Wang HZ, Guo LJ. OptoGPT: a foundation model for inverse design in optical multilayer thin film structures. *Opto-Electron Adv* **7**, 240062 (2024).
- Lee J, Vučković J. Nanophotonic computational design. *Opt Express* **21**, 13351–13367 (2013).
- Piggott AY, Lu J, Lagoudakis KG et al. Inverse design and demonstration of a compact and broadband on-chip wavelength demultiplexer. *Nat Photonics* **9**, 374–377 (2015).
- Pendry JB, Schurig D, Smith DR. Controlling electromagnetic fields. *Science* **312**, 1780–1782 (2006).
- Smith DR, Pendry JB, Wiltshire MCK. Metamaterials and negative refractive index. *Science* **305**, 788–792 (2004).
- Pendry JB. Negative refraction makes a perfect lens. *Phys Rev Lett* **85**, 3966–3969 (2000).
- Chen J, Hu SS, Zhu SN et al. Metamaterials: from fundamental physics to intelligent design. *Interdiscip Mater* **2**, 5–29 (2023).
- Padilla WJ, Averitt RD. Imaging with metamaterials. *Nat Rev Phys* **4**, 85–100 (2022).
- Shalaev MI, Walasik W, Tsukernik A et al. Robust topologically protected transport in photonic crystals at telecommunication wavelengths. *Nat Nanotech* **14**, 31–34 (2019).
- Xiao XF, Gillibert R, Foti A et al. Plasmonic polarization rotation in SERS spectroscopy. *Nano Lett* **23**, 2530–2535 (2023).
- Lee YU, Zhao JX, Ma Q et al. Metamaterial assisted illumination nanoscopy via random super-resolution speckles. *Nat Commun* **12**, 1559 (2021).
- Liu ZW, Lee H, Xiong Y et al. Far-field optical hyperlens magnifying sub-diffraction-limited objects. *Science* **315**, 1686 (2007).
- Tang GJ, He XT, Shi FL et al. Topological photonic crystals: physics, designs, and applications. *Laser Photonics Rev* **16**, 2100300 (2022).
- Tsakmakidis KL, Reshef O, Almpanis E et al. Ultrabroadband 3D invisibility with fast-light cloaks. *Nat Commun* **10**, 4859 (2019).
- Riganti R, Zhu YL, Cai W et al. Multiscale physics-informed neural networks for the inverse design of hyperuniform optical materials. *Adv Opt Mater* **13**, 2403304 (2025).
- Zhang ER, Dao M, Karniadakis GE et al. Analyses of internal structures and defects in materials using physics-informed neural networks. *Sci Adv* **8**, eabk0644 (2022).
- Gu T, Kim HJ, Rivero-Baleine C et al. Reconfigurable metasurfaces towards commercial success. *Nat Photonics* **17**, 48–58 (2023).
- Tikhonravov AV, Trubetskov MK, DeBell GW. Application of the needle optimization technique to the design of optical coatings. *Appl Opt* **35**, 5493–5508 (1996).
- Dobrowolski JA. Numerical methods for optical thin films. *Opt Photonics News* **8**, 24 (1997).
- Holland JH. Outline for a logical theory of adaptive systems. *J ACM* **9**, 297–314 (1962).
- Kennedy J, Eberhart R. Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks 1942–1948* (IEEE, 1995). <http://doi.org/10.1109/ICNN.1995.488968>.
- Martin S, Rivory J, Schoenauer M. Synthesis of optical multilayer systems using genetic algorithms. *Appl Opt* **34**, 2247–2254 (1995).
- Liu ZC, Zhu DY, Rodrigues SP et al. Generative model for the inverse design of metasurfaces. *Nano Lett* **18**, 6570–6576 (2018).
- Ma W, Cheng F, Xu YH et al. Probabilistic representation and inverse design of metamaterials based on a deep generative model with semi-supervised learning strategy. *Adv Mater* **31**, 1901111 (2019).
- Naseri P, Hum SV. A generative machine learning-based approach for inverse design of multilayer metasurfaces. *IEEE Trans Antennas Propag* **69**, 5725–5739 (2021).
- Goodfellow I, Pouget-Abadie J, Mirza M et al. Generative adversarial networks. *Commun ACM* **63**, 139–144 (2020).
- Sutton RS. Learning to predict by the methods of temporal differences. *Mach Learn* **3**, 9–44 (1988).
- Silver D, Huang A, Maddison CJ et al. Mastering the game of Go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016).
- Vaswani A, Shazeer N, Parmar N et al. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* 6000–6010 (Curran Associates Inc., 2017).
- Raissi M, Perdikaris P, Karniadakis GE. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J Comput Phys* **378**, 686–707 (2019).
- Shafiq M, Gu ZQ. Deep residual learning for image recognition: a survey. *Appl Sci* **12**, 8972 (2022).
- Chen W, Gao Y, Li YY et al. Broadband solar metamaterial absorbers empowered by transformer-based deep learning. *Adv Sci* **10**, 2206718 (2023).
- Huang YK, Liu XS, Liu ML et al. Deep learning-based inverse design of multi-functional metasurface absorbers. *Opt Lett* **49**, 2733–2736 (2024).
- Panda SS, Choudhary S, Joshi S et al. Deep learning approach for inverse design of metasurfaces with a wider shape gamut. *Opt Lett* **47**, 2586–2589 (2022).
- Unni R, Yao K, Zheng YB. Deep convolutional mixture density network for inverse design of layered photonic structures. *ACS Photonics* **7**, 2703–2712 (2020).
- Roberts NB, Hedayati MK. A deep learning approach to the forward prediction and inverse design of plasmonic metasurface structural color. *Appl Phys Lett* **119**, 061101 (2021).
- Qian K, Zhang YY. Tunable grating surfaces with high diffractive efficiency optimized by deep neural networks. *Opt Lett* **47**, 4660–4663 (2022).
- Si LM, Niu R, Dang CY et al. Advances in artificial intelligence for artificial metamaterials. *APL Mater* **12**, 120602 (2024).

42. Liu W, Xu GX, Fan W et al. Machine learning-based characterization and inverse design of metamaterials. *Materials* **17**, 3512 (2024).
43. Tezsezen E, Yigci D, Ahmadvpour A et al. AI-based metamaterial design. *ACS Appl Mater Interfaces* **16**, 29547–29569 (2024).
44. Li ZZ, Guo CR, Lv WL et al. Machine learning-enabled optical architecture design of perovskite solar cells. *J Phys Chem Lett* **15**, 3835–3842 (2024).
45. Ma W, Liu ZC, Kudyshev ZA et al. Deep learning for the design of photonic structures. *Nat Photonics* **15**, 77–90 (2021).
46. Qian C, Kaminer I, Chen HS. A guidance to intelligent metamaterials and metamaterials intelligence. *Nat Commun* **16**, 1154 (2025).
47. Wiecha PR, Arbouet A, Girard C et al. Deep learning in nano-photonics: inverse design and beyond. *Photonics Res* **9**, B182–B200 (2021).
48. Lee S, Park C, Rho J. Mapping information and light: trends of AI-enabled metaphotonics. *Curr Opin Solid State Mater Sci* **29**, 101144 (2024).
49. Wang HZ, Guo LJ. NEUTRON: neural particle swarm optimization for material-aware inverse design of structural color. *iScience* **25**, 104339 (2022).
50. Wang HZ, Zheng ZY, Ji CG et al. Automated multi-layer optical design via deep reinforcement learning. *Mach Learn Sci Technol* **2**, 025013 (2021).
51. Saha A, Ma TG, Wang HZ et al. Environmentally sustainable and multifunctional chrome-like coatings having no chromium designed with reinforcement learning. *ACS Appl Mater Interfaces* **15**, 28772–28780 (2023).
52. Jiang AQ, Osamu Y, Chen LY. Multilayer optical thin film design with deep Q learning. *Sci Rep* **10**, 12780 (2020).
53. Reuna J, Polojärvi V, Pääkkönen P et al. Influence of ex-situ annealing on the properties of MgF₂ thin films deposited by electron beam evaporation. *Opt Mater* **96**, 109326 (2019).
54. He LJ, Li C, Liu XZ. The electrical and morphological properties of magnesium oxide/alumina bilayered thin films prepared by electron beam evaporation at oblique incidence. *Appl Surf Sci* **292**, 665–669 (2014).
55. Wu HY, Chen HW, Huang SR et al. Highly reflective silver mirror enhanced by several dielectric films prepared under the low substrate temperature. *Heliyon* **10**, e35507 (2024).
56. Friz M, Waibel F. Coating materials. In Kaiser N, Pulker HK. *Optical Interference Coatings* 105–130 (Springer, Berlin, Heidelberg, 2003).
57. Protopapa ML, De Tomasi F, Perrone MR et al. Laser damage studies on MgF₂ thin films. *J Vac Sci Technol A* **19**, 681–688 (2001).
58. Mihama K, Iwama S, Tanaka N. Growth of crystalline composite films of Au-MoS₂ and CaF₂-MgO. *Z Phys D* **26**, 243–245 (1993).
59. Doğan Y, Erdoğan İ. Antireflection coating for MWIR on calcium fluoride using ion-assisted e-beam deposition. *Celal Bayar Univ J Sci* **21**, 95 (2025).
60. Haering RR, Miksic MG, Pennebaker WB. Method of fabricating thin film transistor devices: US3303067A[P]. 1967-02-07.
61. Song B, Zhao LL, Chen XY et al. Effects of HfO₂ interlayer on microstructure and mechanical property of Al₂O₃ thin film on MgF₂ substrate. *J Inorg Mater* **31**, 779–784 (2016).
62. Skriabin AS, Telekh VD, Pavlov AV et al. Visualization of the gas flows that formed above the thin-film coatings under VUV radiation influence. *J Phys Conf Ser* **2127**, 012005 (2021).
63. Mao SD. Effect of post-treatment process on the properties of optical films [D]. Changchun: Changchun University of Science and Technology, 2020.
64. Bennett HE, Guenther AH, Milam D, et al. Ion Assisted Deposition of Optical Films on Heavy Metal Fluoride Glass Substrates. Natl. Bur. Stand. (U.S.), Spec. Publ. 746, 342 (1988).
65. Syed WAA, Rafiq N, Ali A et al. Multilayer AR coatings of TiO₂/MgF₂ for application in optoelectronic devices. *Optik* **136**, 564–572 (2017).
66. Werner U, Bartsch W, Lindner R, et al. Heat-reflecting welding protection filter: GB2103661B [P]. 1983-02-23.
67. Denny YR, Siswahyu M, Hamaguchi S et al. Effects of oxygen partial pressure on the electronic properties and chemical state of MgO/SiO₂/Si(100) thin films. *Key Eng Mater* **965**, 85–90 (2023).
68. Heitmann W. Properties of evaporated SiO₂, SiO_xN_y, and TiO₂ films. *Appl Opt* **10**, 2685–2689 (1971).
69. Bian LH, Wang Z, Zhang YZ et al. A broadband hyperspectral image sensor with high spatio-temporal resolution. *Nature* **635**, 73–81 (2024).
70. Bao J, Bawendi MG. A colloidal quantum dot spectrometer. *Nature* **523**, 67–70 (2015).
71. Yang ZY, Albrow-Owen T, Cui HX et al. Single-nanowire spectrometers. *Science* **365**, 1017–1020 (2019).

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant Nos. 92570115 and 12074037) for the support of this work.

Author contributions

Y. Zhang conceived the research framework. H. Wu and Y. Zhang carried out the research work, developed architecture and performed the data analysis. B.-C. Li, S. Yuan, J. Peng, and H. Zhong provided constructive suggestions and feedback. The text was written by H. Wu, H. Zhong, and Y. Zhang.

Competing interests

The authors declare no competing financial interests.

Supplementary information

Supplementary information for this paper is available at <https://doi.org/10.67704/ioe.2026.250018>



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>
©The Author(s) 2026.

